

## **RAW SEQUENCE LISTING**

**The Biotechnology Systems Branch of the Scientific and Technical  
Information Center (STIC) no errors detected.**

Application Serial Number: 09/709,020A  
Source: IFWO  
Date Processed by STIC: 04/11/2007

# ***ENTERED***



IFWO

## RAW SEQUENCE LISTING

DATE: 04/11/2007

PATENT APPLICATION: US/09/709,020A

TIME: 09:56:48

Input Set : A:\ES.txt

Output Set: N:\CRF4\04102007\I709020A.raw

```

3 <110> APPLICANT: Benning, Christoph
4      Sanda, Sherrie
5      Yu, Bin
7 <120> TITLE OF INVENTION: Compositions and Methods for the Synthesis and Subsequent
Modification of
8      Uridine-5-Diphosphosulfoquinovose (UDP-SQ)
10 <130> FILE REFERENCE: MSU-04769
C--> 12 <140> CURRENT APPLICATION NUMBER: US/09/709,020A
C--> 12 <141> CURRENT FILING DATE: 2000-11-08
12 <160> NUMBER OF SEQ ID NOS: 24
14 <170> SOFTWARE: PatentIn version 3.0
16 <210> SEQ ID NO: 1
17 <211> LENGTH: 1134
18 <212> TYPE: DNA
19 <213> ORGANISM: Cyanobacterium Synechococcus
21 <220> FEATURE:
22 <221> NAME/KEY: CDS
23 <222> LOCATION: (1)..(1134)
25 <400> SEQUENCE: 1
26 atg cgc atc gct ctc ttt acc gag acg ttc ctc ccc aaa gtg gat ggc      48
27 Met Arg Ile Ala Leu Phe Thr Glu Thr Phe Leu Pro Lys Val Asp Gly
28 1          5          10          15
30 atc gtc acg cgg ctt cgg cac acg gtc gat cac ctg cag cgt ctt ggc      96
31 Ile Val Thr Arg Leu Arg His Thr Val Asp His Leu Gln Arg Leu Gly
32          20          25          30
34 cac acc gtc atg gtt ttt tgc ccc gac ggc ggg ctc cgc gag cac aag      144
35 His Thr Val Met Val Phe Cys Pro Asp Gly Gly Leu Arg Glu His Lys
36          35          40          45
38 ggg gct cga gtc tat ggg gtt aaa ggc ttt ccg cta ccg ctc tat ccc      192
39 Gly Ala Arg Val Tyr Gly Val Lys Gly Phe Pro Leu Pro Leu Tyr Pro
40          50          55          60
42 gag ctg aag cta gct ttt ccg ttg ccg aaa gtg gga aaa gcc ttg gag      240
43 Glu Leu Lys Leu Ala Phe Pro Leu Pro Lys Val Gly Lys Ala Leu Glu
44 65          70          75          80
46 cgg ttc cgg ccc gac ctg atc cac gtg gtc aat ccg gct gtg ttg ggg      288
47 Arg Phe Arg Pro Asp Leu Ile His Val Val Asn Pro Ala Val Leu Gly
48          85          90          95
50 ttg ggc ggc atc tac tat gcc aag gcg cta aat gtg cca ctc gtg gcg      336
51 Leu Gly Gly Ile Tyr Tyr Ala Lys Ala Leu Asn Val Pro Leu Val Ala
52          100         105         110
54 tcc tat cac acc cat ttg ccg aaa tac ctt gag cat tac ggg ctg ggg      384
55 Ser Tyr His Thr His Leu Pro Lys Tyr Leu Glu His Tyr Gly Leu Gly
56          115         120         125
58 gtc ttg gag ggg gtg ctc tgg gaa ttg ctg aag ctg gcg cat aac caa      432

```

## RAW SEQUENCE LISTING

DATE: 04/11/2007

PATENT APPLICATION: US/09/709,020A

TIME: 09:56:48

Input Set : A:\ES.txt

Output Set: N:\CRF4\04102007\I709020A.raw

```

59 Val Leu Glu Gly Val Leu Trp Glu Leu Leu Lys Leu Ala His Asn Gln
60      130      135      140
62 gca gcg atc aac ctc tgt act tca acc gcg atg gtg cag gag ctg aca      480
63 Ala Ala Ile Asn Leu Cys Thr Ser Thr Ala Met Val Gln Glu Leu Thr
64 145      150      155      160
66 gat cac ggc att gag cac tgt tgc ctc tgg cag cga gga gtg gat acc      528
67 Asp His Gly Ile Glu His Cys Cys Leu Trp Gln Arg Gly Val Asp Thr
68      165      170      175
70 gag acc ttt cgg cca gac ttg gct act gct gcg atg cgc gat cgc ctc      576
71 Glu Thr Phe Arg Pro Asp Leu Ala Thr Ala Ala Met Arg Asp Arg Leu
72      180      185      190
74 agt ggc ggt aag ccc act gcg ccc ttg ttg ctc tac gtc gga cgc ctc      624
75 Ser Gly Gly Lys Pro Thr Ala Pro Leu Leu Leu Tyr Val Gly Arg Leu
76      195      200      205
78 tca gcc gag aag caa atc gat cgc ctg cga ccc att ttg gat gcc aat      672
79 Ser Ala Glu Lys Gln Ile Asp Arg Leu Arg Pro Ile Leu Asp Ala Asn
80      210      215      220
82 cct gag gct tgc ttg gcc ttg gtc ggc gat ggc ccg cat cgg gcc gaa      720
83 Pro Glu Ala Cys Leu Ala Leu Val Gly Asp Gly Pro His Arg Ala Glu
84 225      230      235      240
86 cta gag caa ttg ttt gct ggc acc cag acg cag ttc att ggc tat ctg      768
87 Leu Glu Gln Leu Phe Ala Gly Thr Gln Thr Gln Phe Ile Gly Tyr Leu
88      245      250      255
90 cat ggg gaa cag cta ggg gcg gcc tac gct tct gct gac gct ttt gtc      816
91 His Gly Glu Gln Leu Gly Ala Ala Tyr Ala Ser Ala Asp Ala Phe Val
92      260      265      270
94 ttt ccc tcc cgg acc gaa acc ctc ggt cta gtc ttg ctg gaa gcc atg      864
95 Phe Pro Ser Arg Thr Glu Thr Leu Gly Leu Val Leu Leu Glu Ala Met
96      275      280      285
98 gca gcg ggt tgt ccg gtc gtg gcg gcc aat tcc ggt ggc att ccc gat      912
99 Ala Ala Gly Cys Pro Val Val Ala Ala Asn Ser Gly Gly Ile Pro Asp
100      290      295      300
102 att gtc agc gac ggc att aat ggt ttc ctg ttc gat cct gag gat gaa      960
103 Ile Val Ser Asp Gly Ile Asn Gly Phe Leu Phe Asp Pro Glu Asp Glu
104 305      310      315      320
106 caa ggg gcg atc gct gcg att cag cgc ttg ttg gct aac cct gca gag      1008
107 Gln Gly Ala Ile Ala Ala Ile Gln Arg Leu Leu Ala Asn Pro Ala Glu
108      325      330      335
110 cgc gag att cta cgc caa gcg gct cgt caa gaa gcc gaa cgc tgg agc      1056
111 Arg Glu Ile Leu Arg Gln Ala Ala Arg Gln Glu Ala Glu Arg Trp Ser
112      340      345      350
114 tgg aac gca gcc acg cgc caa ctc cag gac tac tac tgc gag gtg ttg      1104
115 Trp Asn Ala Ala Thr Arg Gln Leu Gln Asp Tyr Tyr Cys Glu Val Leu
116      355      360      365
118 gca gat ggt tgc tta ccc tta gcg gcc tga      1134
119 Ala Asp Gly Cys Leu Pro Leu Ala Ala
120      370      375
123 <210> SEQ ID NO: 2
124 <211> LENGTH: 377

```

## RAW SEQUENCE LISTING

DATE: 04/11/2007

PATENT APPLICATION: US/09/709,020A

TIME: 09:56:48

Input Set : A:\ES.txt

Output Set: N:\CRF4\04102007\I709020A.raw

```

125 <212> TYPE: PRT
126 <213> ORGANISM: Cyanobacterium Synechococcus
128 <400> SEQUENCE: 2
130 Met Arg Ile Ala Leu Phe Thr Glu Thr Phe Leu Pro Lys Val Asp Gly
131 1 5 10 15
134 Ile Val Thr Arg Leu Arg His Thr Val Asp His Leu Gln Arg Leu Gly
135 20 25 30
138 His Thr Val Met Val Phe Cys Pro Asp Gly Gly Leu Arg Glu His Lys
139 35 40 45
142 Gly Ala Arg Val Tyr Gly Val Lys Gly Phe Pro Leu Pro Leu Tyr Pro
143 50 55 60
146 Glu Leu Lys Leu Ala Phe Pro Leu Pro Lys Val Gly Lys Ala Leu Glu
147 65 70 75 80
150 Arg Phe Arg Pro Asp Leu Ile His Val Val Asn Pro Ala Val Leu Gly
151 85 90 95
154 Leu Gly Gly Ile Tyr Tyr Ala Lys Ala Leu Asn Val Pro Leu Val Ala
155 100 105 110
158 Ser Tyr His Thr His Leu Pro Lys Tyr Leu Glu His Tyr Gly Leu Gly
159 115 120 125
162 Val Leu Glu Gly Val Leu Trp Glu Leu Leu Lys Leu Ala His Asn Gln
163 130 135 140
166 Ala Ala Ile Asn Leu Cys Thr Ser Thr Ala Met Val Gln Glu Leu Thr
167 145 150 155 160
170 Asp His Gly Ile Glu His Cys Cys Leu Trp Gln Arg Gly Val Asp Thr
171 165 170 175
174 Glu Thr Phe Arg Pro Asp Leu Ala Thr Ala Ala Met Arg Asp Arg Leu
175 180 185 190
178 Ser Gly Gly Lys Pro Thr Ala Pro Leu Leu Leu Tyr Val Gly Arg Leu
179 195 200 205
182 Ser Ala Glu Lys Gln Ile Asp Arg Leu Arg Pro Ile Leu Asp Ala Asn
183 210 215 220
186 Pro Glu Ala Cys Leu Ala Leu Val Gly Asp Gly Pro His Arg Ala Glu
187 225 230 235 240
190 Leu Glu Gln Leu Phe Ala Gly Thr Gln Thr Gln Phe Ile Gly Tyr Leu
191 245 250 255
194 His Gly Glu Gln Leu Gly Ala Ala Tyr Ala Ser Ala Asp Ala Phe Val
195 260 265 270
198 Phe Pro Ser Arg Thr Glu Thr Leu Gly Leu Val Leu Leu Glu Ala Met
199 275 280 285
202 Ala Ala Gly Cys Pro Val Val Ala Ala Asn Ser Gly Gly Ile Pro Asp
203 290 295 300
206 Ile Val Ser Asp Gly Ile Asn Gly Phe Leu Phe Asp Pro Glu Asp Glu
207 305 310 315 320
210 Gln Gly Ala Ile Ala Ala Ile Gln Arg Leu Leu Ala Asn Pro Ala Glu
211 325 330 335
214 Arg Glu Ile Leu Arg Gln Ala Ala Arg Gln Glu Ala Glu Arg Trp Ser
215 340 345 350
218 Trp Asn Ala Ala Thr Arg Gln Leu Gln Asp Tyr Tyr Cys Glu Val Leu
219 355 360 365

```

## RAW SEQUENCE LISTING

DATE: 04/11/2007

PATENT APPLICATION: US/09/709,020A

TIME: 09:56:48

Input Set : A:\ES.txt

Output Set: N:\CRF4\04102007\I709020A.raw

222 Ala Asp Gly Cys Leu Pro Leu Ala Ala

223 370 375

226 &lt;210&gt; SEQ ID NO: 3

227 &lt;211&gt; LENGTH: 2979

228 &lt;212&gt; TYPE: DNA

229 &lt;213&gt; ORGANISM: Cyanobacterium Synechococcus

231 &lt;400&gt; SEQUENCE: 3

```

232 ctacacgtta ccttcgggta ctggaaacag tcgtttaatc aaccaattga ttgggtcccaa      60
234 aacatgaact ttctttttcc tccagaacca aatcgctgca ctgtactgtt cattgcgtat      120
236 ctttgtcgtt gctgctctcc aatcatatct ctcgggtctct tctcttgccg cttttccaat      180
238 gatctctctt gtttcacggg cgtgcagtaa agttctcagt tttgtcacgc aatcttcaac      240
240 atctccaggg ttgaacaaaa atccggtttt tccctcctga tgaaaacatc agaatcagaa      300
242 aaccacaagc tcaatatagg ttgaccata agaacaatca atgcaagatc attttgtgta      360
244 ccagtcctat attgaataaa gtttcagttc ggttacagct cgcttataag aaaattggca      420
246 gaaattgttt tttcaacctt ttccggttcg ttgatatgct catcaatatg gtttggcagt      480
248 taattgtaat tcagataaatt cactgacctg atcttcaggg attatatcag ggattccacc      540
250 ggcacggggc gcgacgacag gaagtcctga agacattgct tcaagaacca caaggccaag      600
252 tgtctccgac tctgatggca tcacaaacac atctccactt gcgtaagctt gtgagagttc      660
254 atcgccctgt aacgttccag tgaaaaccgc tggcattccg gtaaacactt tctcaagatc      720
256 ctctctttta gaaaacgaaa cagataaaca aaattacaat gttgttgact agaaatcttc      780
258 agataacaat atggccaatc tttaacaaaa ctagtacttg tatgggtccat ctccaatgaa      840
260 agcaatccga gcttcaggta atttgtccat tacactgcac acaaatttct caatatcaaa      900
262 attcgataca ccacttaaaa gaagtgaagc cagtttatac aaaattctaa cctctttaaa      960
264 agctccaaac tcttttctac gccaatgcga cctacatgaa tcaactagtg cttttctggt      1020
266 tcgccattac tgtaattac aaaatattaa acatcaagat tagcgtggaa agtatcattg      1080
268 tttttaatgc atataaaaaga aacgtatatt ctattcttgc ctcagtctta tacgcatttc      1140
270 ttgagaacgg aaacggggat tgaagctttc tgaatcgaca cccttattcc aaagtcgaag      1200
272 ttgattagct tcaatatatg aaaaaaagaa gaagaaaatg taagttttga acaatcatag      1260
274 agcttgtaaa aaaatgtaat gtatgatctt tcttcacctg cagttgcacc agctgctata      1320
276 agatcttttc caatggcagc agaaggaact aatgtaagat cagccgctct gtgaaggaac      1380
278 cctgataaaa gcatattcaa gtttagtttc atattataca tcacaataaa ccagaaaaag      1440
280 aagaaggaag ttttgacatt tgaaaagcgg gttttacata cttattatag accacattgg      1500
282 ttttaccacaa caactaaaag tgtatcttgg gatgtatctg cacatcaaca caagcttaat      1560
284 cttagacaaa aattttttta taacaacatt gtgaaatgag gcagaaaaag gtacttacac      1620
286 agggacgtgt gtgtggtaag acattactat tggtagagat agcatttttg ctattgccag      1680
288 agcaccaaag acctaaaaat tttagtcagg gaaaaaagag agtcaagttc tggattctct      1740
290 ccagttcact ggtcttgcta atgttttagt aatgtgaatt cttgagggat ttaccataac      1800
292 tccggggagat gaagcgtgta taatgtcagg cttaaaccgt gcaatttcag agatgattct      1860
294 tggactaagc gcaagcgaga gtggaacctt ttggtaataa ggacaaggga agctacagaa      1920
296 gagaagaaga aattagcgat attacaaat agagaacatc cagtgaagta actaaatggg      1980
298 gctaccttct tgatccaatg actctggctc cataaaactc ttcaggaaca cttcatgtg      2040
300 tcgtcacgac tataacctat gaagcaaaaa agttattaaa aaaaaaaaaa aaggaaacag      2100
302 taacacttgt caagtaattc taattctgga aacagttact tatgagctga ctgaaaagat      2160
304 acttaagtgt aagaatgaga tagtaaaaaga agaaacctcg tctccattt cacggaggta      2220
306 tctaatgaaa ttctggaatc tgtttttgta gccggataca tagctgcaac aaaatcaaag      2280
308 agagaatcac ttccaataat aacatgacat ataataaaag cttttgggtca atggatcggg      2340
310 gattccgaga atcttgggat attcacaact aaaatctgac aactttgact caaacaaatc      2400
312 ctgaatgtaa ttggttttta cgatctacta tataatttgc taaattgggt gtgtagcaaa      2460
314 ttcatacatt agcgagtatc ttttcataaa ataaatgtaa cgatcaaac gaaagaaaaa      2520

```

## RAW SEQUENCE LISTING

DATE: 04/11/2007

PATENT APPLICATION: US/09/709,020A

TIME: 09:56:48

Input Set : A:\ES.txt

Output Set: N:\CRF4\04102007\I709020A.raw

```

316 aacattacag gaaaaagtca accaaggaaa aaaatgagta gaatctgttt tcacagagac 2580
318 atttcgtcga acacaaaaca agcaaaaaaa gaacactgtg aagaagactt acgcaaaggg 2640
320 agaaggctca acaaagagag caattctcct aggcttagag agcgactcag gatcgagaag 2700
322 cggcgcatca atctccgatt cgtcatcttc tctgacttga gtaatagtca tatcgttgga 2760
324 cccagaaaaca gcttcctttg taatcactcc acagaagcgg agcttgcttt tcttgctaata 2820
326 gggtaatctc cggtgaccaa acgaaagagg ggaatgaaga acaaaagaag aagacctggg 2880
328 aggagaacaa gaggttgacg aggaagaaga acaagtgtta gtcgtgctag gaagcaaata 2940
330 aggaggtata gagagattta tagaagaaag agtcgtcat 2979
333 <210> SEQ ID NO: 4
334 <211> LENGTH: 2537
335 <212> TYPE: DNA
336 <213> ORGANISM: Arabidopsis thaliana
338 <400> SEQUENCE: 4
339 tcatattttg aaaaagcact tcttcttctt cttcttcttc ttcttcttct tcttcttccc 60
341 atcaaaaagt gctagtggca agaccgcgca tattccagta acaaagttta aattcaaaac 120
343 tatggaacac aatttgtccc atgttttggt gcctgaatta aacgaccata tcacttggtt 180
345 cggccacttc atctcggata cgcacaagcg gttgtcaagg ttgtaaaaga cgattttgta 240
347 aggatctaca ttggcaaagg gagctttaca gacgacttga aaagtctcgg tgtgaagatt 300
349 gaaaggtaaa actttggttt ctttgcactc ggtgaaccaa taaagcgacc catctagata 360
351 cacagggtcg gggtaagcag caatccgata aggagcagcg agagtgcacat acctccaagc 420
353 attggtacta aagtcgaaaa ctctcgatgt agtagcgttt tttaggccta tttctgcaga 480
355 gttgtataac caaacgggct tgtatgtgcc cctgaatttg tctttaccga atccaagcat 540
357 aaagaatttg cgtttaagct tgtagtaacc atctcgtaag tcgatcatga gtttttgata 600
359 atcgcaaaga ggaagaggcc gataccatct agtggtcggg ttgaccacat aaccggattt 660
361 gtcgtgattg taaagacaaa cgagaccgtc acaactacta tgtgaaacta agtacagtac 720
363 gttatccttc tcccaaggag tagggatctt gaacgttgat gatgaacca actgcaatgt 780
365 tcttaaagat tctatggctg gggtttataac atgggtgagga aatacagaca ccatcagaac 840
367 atctggatct cctgattgct gacgatgttt caactaaaat tgtatacatt tttattttat 900
369 taaaaagata aatatcattt tgttgggaatt ggaaaacaaa atctaaaaga taactttaga 960
371 tatgtaaggg tcgcatatgt atttgattac attgctatga gtatatgact cacttttcca 1020
373 ataataaaaa aaataaagat gttgagctct actaattaag ggtaaatata aaatttctga 1080
375 tcattaatac aaagaaaaag cctctcataa gccaaagcca ttttactcgc cgtaaaccatc 1140
377 tccgcagcat actctttaca cttccttctc ctctccgcca atctctccgc tccttcagcc 1200
379 acagcaacct ccataaccgc cgtcaaagct tccacatttg gcgcaaacad aaacccaaac 1260
381 tcatcattca ccactatagt cctctttatg cttgcgtacc tagatgccat cactggttta 1320
383 ccactcaaca tcgcttccat taacgttaaa tcaagacctt gtggtctaag cgttggattc 1380
385 acgaacaaat cgatcccggt gtagaaaccc ttgagctcat ttgggtttag agatcccaa 1440
387 atggaaactt tttcccctaa ttccttgtaa cgttgctccc atggccctga tccagctact 1500
389 acgaggtaaa catttgaaata cgtttggtatt attttcgcga aagcttcgaa gagcaatgga 1560
391 tgtcctttgt ctttgactaa tctcccagca gtccttaaaa caatcgctga tgagttttct 1620
393 ggtaacccta attttgacct aaacagagta cgtagcttct tgtctgatgt gaatccgttc 1680
395 tcgtcgactc cattgaggat cacatgaacc cttttctcag ggatttggtta aacgtctctt 1740
397 agcatttctc cgcagctatc gctgatagcg atgtggtgag cgtagtgttg gaagaatctg 1800
399 atttcgtcca gtatcttggg aagcacggca ccgtatagac ttgcattgaa gccttgatgat 1860
401 cttgggttcgt ctgggtttacg gatcagggtct tggtaaaatac ttgattgtaa gctctctaac 1920
403 gcaatgccgt gccaggatac agcgagggtt ggaacctccc gggcgatcca gtgaggtaaa 1980
405 gcaaacactt cagagtgaac cgcacgaaa gggtctttct tgttttcttc ttggtaaagc 2040
407 tcccatgcct tgttgtagcg ccattttcgg ggctccgcgt cgccgtgaga atggattata 2100
409 gggtaaatga tttggtcggg aacagggggg attttgttgg tttcagggga ttggtctaag 2160

```

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/09/709,020A

DATE: 04/11/2007  
TIME: 09:56:49

Input Set : A:\ES.txt

Output Set: N:\CRF4\04102007\I709020A.raw

Invalid Line Length:

The rules require that a line not exceed 72 characters in length. This includes spaces.

Seq#:1; Line(s) 7

**VERIFICATION SUMMARY**

DATE: 04/11/2007

PATENT APPLICATION: US/09/709,020A

TIME: 09:56:49

Input Set : A:\ES.txt

Output Set: N:\CRF4\04102007\I709020A.raw

L:12 M:270 C: Current Application Number differs, Replaced Current Application No

L:12 M:271 C: Current Filing Date differs, Replaced Current Filing Date